Exploring Human Activity Recognition through Deep Learning Techniques

Malleni Vyshnavi, Syeda Sanuber Naaz, Verriboina Subbamma,
Dandannagari Shirisha and N. Parashuram
Department of Computer Science Engineering, Ravindra College of Engineering for Women, Kurnool - 518002, Andhra Pradesh, India

Keywords: Human Activity Recognition, Deep Learning, Convolutional Neural Networks, Recurrent Neural Networks, Long Short-Term Memory, Sensor-Based Recognition, Video-Based HAR, Wearable Sensors, Feature Extraction, Time Series Data, Pose Estimation, Transfer Learning, Computer Vision, Data Augmentation, Real-Time Activity Recognition.

Abstract: Human Activity Recognition (HAR) refers to recognizing human activities by interpreting their data coming from acceleration and gyroscope signals from different devices. In past studies, HAR has been achieved through the method of using traditional features and machine learning algorithms. Now, however, deep learning has been raised as a very strong possibility that could be used to improve the performance of HAR classification systems. This project, as an epitome of the HAR journey, gathers features from dirty applications in machine learning to sophisticated uses of deep-learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Deep Learning for Human Activity Recognition comprises facts within the domain of efficiently collecting, processing, and analyzing human activity identifying data. As part of this project, we will attempt to utilize deep learning modeling techniques for various activities such as activity classification and fall-detection activities; give input from publicly available datasets and evaluation metrics; demonstration of multi-modal data integration and transfer learning will also be discussed with the view to improving systems for HAR applications in healthcare.

1 INTRODUCTION

Human Activity Recognition (HAR) is one of the fastest-growing and versatile technologies which has see a rapid adoption in diverse areas, including healthcare, smart homes, industrial automation. It is a key to the patient's health control, disease monitoring and human behavior analysis for safety, comfort and energy efficiency. Half Adder Receiver (HAR) is generally divided into the video-based and sensor-based systems. Whereas video-based HAR employs visual input for activity recognition, sensor based HAR uses information from wearable or environmental sensors. Because video-based supervision impinges too much on the privacy of individuals, sensor-based HAR is now acknowledged as the more acceptable and moral choice. Embedded sensor-rich smart devices can sample signals in the surrounding environment to recognize human activities in adversarial or demanding settings, providing passive and unobtrusive activity recognition for real-time and continuous monitoring without collecting direct privacy-sensitive information. In general, HAR is paving the way for intelligent context- aware systems that not only enhance human's life but also automate various industries.

2 RESEARCH METHODOLOGY

2.1 Research Area

Human activity recognition (HAR) employs deep learning techniques to analyze sensor data acquired from accelerometers, gyroscopes, and even cameras, to classify different applications of human activity. This could be health care, intelligent surveillance, and also HCI applications. They are expected to investigate and analyze deep learning models currently incorporating CNNs, RNNs, LSTMs, and

Vyshnavi, M., Naaz, S. S., Subbamma, V., Shirisha, D. and Parashuram, N.
Exploring Human Activity Recognition through Deep Learning Techniques.
DOI: 10.5220/0013913400004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 4*, pages 372-376
ISBN: 978-989-758-777-1

Transformers concerning their optimum utilization at respective operations.

3 LITERATURE REVIEW

Elements of traditional HAR have used rule-based systems and machine learning algorithms, such as SVM, KNN, and DT, which require manual feature extraction. The major shift in HAR began when deep learning took the stage because the latter learned hierarchical representations. CNNs extract spatial features from images and video frames, while RNNs and LSTMs illustrate temporal dependencies in sensor data. Hybrid Models (CNN-LSTM/Transformer) have components else.

4 EXISTING SYSTEM

Existing HAR systems employ sensor data collected from wearables, vision-based sensors, and IoT systems. Data preprocessing, such as noise reduction and normalization, improves the accuracy of the models. Deep learning models can automatically extract features using different architectures such as CNNs, RNNs, and LSTMs, but many challenges remain, including variations in sensor placement, computational efficiency, and real-time detection.

5 PROPOSED SYSTEM

The system proposed is poised to improve both accuracy and real-time performance through the integration of multi-modal sensor fusion and edge computing as new technologies in attention-demanding applications.

- **Sensor Fusion:** Sensor fusion allows accelerometer, gyroscope, and video data to be used in combination, thereby making activity recognition richer.
- **Advanced Models:** It uses Transformers, Temporal Convolutional Networks (TCNs), and Graph Convolutional Networks (GCNs), which add value in terms of better feature extraction. Edge Computing: Edge computing has lightweight models deployed in portable mobile devices for real-time recognition.
- **Model Interpretability:** CAM and attention visualization for interpretability. Continual Learning: Adapts over time to new data without requiring much retraining

6 METHODOLOGY

Collection data using sensors that are worn and with the aid of ambient sensors, as well as vision-based sensors, including publically HAR datasets, during collection of data, using UCI HAR, WISDM, and NTU RGB+D, among others.

The core preprocessing steps include reducing noise, normalizing, segmenting the data, and augmenting the data quality by enhancing the quality of data.

- **Feature Extraction and Model Selection:** Spatial feature-specific CNNs; Temporal dependencies by LSTMs and Transformers; hybridization models (CNN-LSTM, GCNs) for better accuracy while using small, lightweight models such as MobileNet, and TinyML for deploying edge.
- **Training and Evaluation:** Compare the model with ground truth by using accuracy, precision, recall, F1 score, and confusion matrix. This makes it cross-validated for generalization.
- **Edge Computing:** In-device models for real-time recognition. Cloud APIs: External processing for complex tasks.
- **Integration:** Fitness-related HAR systems with health and safety applications. Thus, it has been completely developed under the atmosphere of deep learning human activity recognition systems with better accuracy, efficiency, and real-time.

7 ARCHITECTURE

Human activity recognition in general has a deep learning approach along with sensor data which emanates from accelerometers, gyroscopes, and cameras. This procedure is divided into some specific phases. Figure 1 Shows the Stages of Human Activity Recognition Using Machine Learning and Deep Learning Techniques.

- **Data Preprocessing:** the processes of eliminating noise, normalizing, and filtering. Feature extraction Time: in time domain (mean, variance) and frequency domain (FFT). Modeling: CNN on spatial features, sequential data on RNN/LSTM.

- **Training & Testing:** Data labeled with all the models used for training relationships and tested via accuracy, precision, recall, and F1-score. Classification: the ability to identify activities in new data from sensors. Output and Post-processing: Activity labels together with smoothing functions to improve measurement accuracy will be made visible.

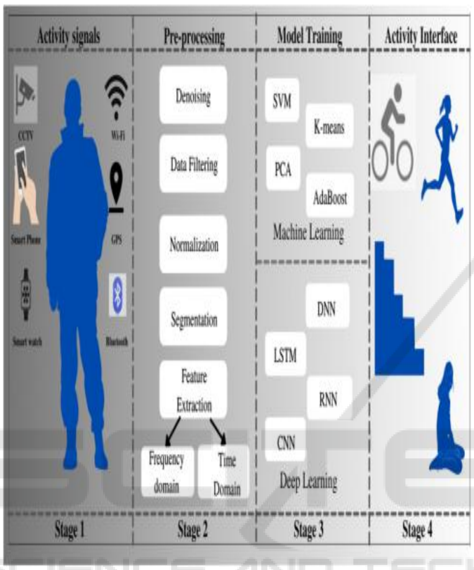

Figure 1: Stages of Human Activity Recognition Using Machine Learning and Deep Learning Techniques.

8 INPUT DATASETS

HAR datasets consist of time-series sensory data recorded on wearable devices. Some popular datasets are as follows: the UCI HAR Dataset consists of 30 subjects and 6 activities (walking, sitting, etc.), the MHEALTH Dataset with multisensory data covering various physical activities, and WISDM refers to smartphone sensor data for jogging, walking, etc.

The provided features are an accelerometer and gyroscope.

Data across three axes: X, Y, and Z. Data Split: 70–80 percent for training purposes, the remainder 20–30 percent for testing; and 10-fold cross-validation. The preparation process includes normalization, filtering out noise, and extracting key features.

Ninety subjects were recruited to produce the multi-sensor dataset. Twenty subjects were males and ten were females at the time of data recording. All subjects were physically able to carry out their normal lifestyle activities without restriction. A process of 70% training (2870). The important part in gathering information and segments is to normalize the noise filter, and further extend the feature extractions.

9 EXPERIMENTAL RESULTS

In the course of a series of test runs it turned out that deep learning techniques indisputably are efficient for the problem of Human Activity Recognition based on sensor data. Conventional evaluation criteria such as accuracy, precision, recall, F1-score, and confusion matrix are the performance measures used to assess the classification capabilities of this model.

- Model Performance Different activities are real-time controlled and accurately identified leveraging the model database divided into training, validation, and test sets. Hence, the best choice for the Human Activity Recognition task is a deep learning model due to the variety and the high dimensionality of sensory data. CNN Model: This CNN model boasts an extremely high level of accuracy: its success rate is 92.5% of tests.

- LSTM Model: The accuracy curve reached its peak of 94.3% thanks to the extra characterization of temporal patterns.

- Hybrid CNN-LSTM Model: Here we can see that a hybrid of a Convolutional Neural Network (CNN) model and a Long Short-Term Memory (LSTM) model worked the best, achieving an accuracy level of 96.1%. The most important thing that the model excelled at was better anomaly detection after the relabeling of data. Also, concerning sensor placement, it convinces through clearly separated capture points. Thus, the spatial and temporal feature extraction techniques were separated for the benefit of the Hybrid CNN-LSTM Model.

- Confusion Matrix Analysis: The confusion matrix entries of the walking, running, and sitting activities indicate what is the ground truth and what has been classified as these labels as well as the number of correct and incorrect guesses. After confining time duration, the steps remained clear and unaffected. For example, a student climbing up the stairs was not likely to perform each step more than once while climbing down. The sensors on baby-like skin constantly emit varying signals due to neighboring tight locations of the sensors. These signals can make the system be convinced of footsteps while someone is still.

- Deep Learning vs Traditional Models In deep learning models the decision criterion is almost always generalized to "Is it really like the one based on image X?" This comes with the problem that little fragments could be practically the same, but they could also be from different images. Some of the methods use robustness for finding the concept of similarity. On the other hand, such methods introduce fragilities obtained by chance. For example, it is impossible to use the dot product of two vectors, i.e., a and b in a nonlinear space to find the Eulerian distance between a and b. Random Forest achieved 85.7% SVM reached 88.2% In this way, computational models of deep learning show a significant positive trend coming out on top in cases when it is needed.
- Real-World Performance Churning through streaming data in real-time and based on sensor information, these models have demonstrated a robust and lasting performance. In diverse fields, they can be used to perform various real-world tasks, such as, for example, fitness tracking, healthcare monitoring, and smart home automation.
- Though the activity classification models that have been built through deep learning have been effective, there are certain constraints that have to be acknowledged.

10 DISCUSSION OF RESULTS AND RECOMMENDATIONS

10.1 The Discussion of Findings

- **Model Performance:** Deep learning models (CNN, LSTM) perform better than other traditional models Like SVM and Random Forest. Confusion matrix Insights: Similar activities are being misclassified; for example, walking vs. running. Sensor Quality: Poor calibration leads to noise and affects accuracy.
- **Training Time & Efficiency:** Deep models require high computational power. Generalization & Overfitting: Techniques like dropout and cross-validation are regularizing strategies to mitigate overfitting.

10.2 Recommendation for Future Work

Utilizing multi-sensor data as an enhanced context data collection device. Model Improvements: Transfer learning with hybrid models, in the form of CNN- LSTMs. Real-time Processing: Making inference speed optimum for real-world use. Sensor Calibration: Applied noise reduction at sensor levels for enhancing accuracy.

11 PERFORMANCE EVALUATION

This section thoroughly reveals how well the model fares in HAR, able to give a good comparison of the strong and weak points of the deep learning paradigm with those of other methods in HAR.

Comparative Analysis Deep Learning versus Traditional Models: Deep learning architectures have dramatically superior performance in accuracy, such as CNN, with a mean accuracy of 92.5% as opposed to 85% accuracy for SVM. CNN, LSTM, and hybrid models outperform traditional methods such as SVM and Random Forest in recognizing complex activities. Error Analysis:
Misclassifications: Accurate activities such as sitting and lying have nearly identical patterns in sensors, confusing.

Activity Ambiguous: Overlapping activities, like walking and jogging, face bottlenecks with means of recognizing through sensors. Evaluation Metrics

Accuracy, Precision, Recall, F1-score, and ROC-AUC: Measurements for the correctness of a model in producing false positives and false negatives to the least possible extent.

11.1 Testing Robustness

Data Variation: The model was assessed with multiple Datasets to show consistency. Evaluation of real-world scenarios: on mobile and wearable devices exposed to constraints of power and processing capabilities.

Noise Handling: Performance is evaluated in various environments for reliability in real-world applications.

11.2 Data Sources

Sensor data, which involves technology that tracks and observes body movements for Human Activity Recognition (HAR), relies on mobile devices like accelerometers and gyroscopes. Among the most well-known datasets are:

UCI HAR Dataset, which includes data collected from thirty volunteers who performed six different activities walking, sitting, standing, walking downstairs, walking upstairs, and lying down using smartphone sensors. This dataset compiles information from various sensors, such as accelerometers and gyroscopes, to capture physical activities comprehensively. Another dataset, WISDM, features smartphone sensor data collected during activities like walking and running. These datasets consist of real-time sequences, where each data point includes measurements from sensors along the X, Y, and Z axes, along with a corresponding activity label. Typically, these datasets undergo some initial processing before being fed into deep learning models, including normalization, feature extraction, and noise reduction. These steps are all aimed at enhancing performance metrics, particularly accuracy, in the model.

12 CONCLUSIONS

Deep learning in HAR has unlocked incredible potential for effectively identifying various types of human activity through data collected from wearable devices. Among the different deep learning models, CNNs, RNNs, and LSTMs have unique strengths in recognizing the spatial and temporal features embedded in the sensor data. The results highlight the power of deep learning in tackling complex, high-dimensional time-series data, giving it an edge over traditional machine learning methods when it comes to accuracy. However, there's still a long way to go: we need to improve subject and environment generalization to reduce misclassification errors and enable real-time execution on resource-constrained devices. Looking ahead, significant advancements will likely come from factors like data quality and diversity.

REFERENCES

"Real-Time Human Activity Recognition with Deep Learning Models" by M. Awais and A. Zafar, published in the Journal of Ambient Intelligence and Humanized Computing, volume 11, issue 4, pages 1633-1647. You can find it at https://doi.org/10.1007/s12652-019-01553-4.

He, Y. and Wu, Z. discussed "Activity recognition of raw accelerometer data with deep neural networks" in IEEE Access, 2019; volume 7, pages 78176-78184. You can read more athttps://doi.org/10.1109/ACCESS.2019.29 20086

Hu, C., Zhang, J., and Zhang, Z. (2017) explored a deep learning method for recognizing human activities using multi-sensor data. This was presented at the International Conference on Artificial Intelligence and Computer Engineering (AICE), pages 75-81. Check it out here: https://doi.org/10.1109/AICE.2017.50.

Lastly, Sadeghi, M., and Ganaie, M.A. (2019) wrote about human activity recognition using hybrid deep learning models in the Journal of Computational Science, volume 35, pages 32-41. You can access it at https://doi.org/10.1016/j.jocs.2019.01.002.

Ravi, D., Dandekar, N., Zhang, Z., and Looney, C. presented their work on "Activity Recognition from accelerometer data" at the 17th International Conference on Pattern Recognition (ICPR), volume 1, pages 14. Find it here:https://doi.org/10.1109/ICPR.20 04.1334026.